The Sound of Deceit: Understanding Voice Cloning Risks

The aim of Disinfo Radar’s Rapid Response Briefs is to identify new and noteworthy disinformation technologies, tactics and narratives. Such cases may include the identification of a new technology that may have such harmful uses.

Background

In 2020, using an AI-generated voice, a group of cybercriminals were able to steal over $35 million from a Japanese company. Mimicking the director of the firm’s parent company, they authorized the company to transfer millions of dollars to a series of dummy accounts spread across the globe. While this incident and others raised concerns about the potential use of voice cloning for fraud, the technology was still nascent, and considered too difficult and expensive to use in all but the most sophisticated cybercrimes.

Today, a mere three years later, internet users are now a simple YouTube tutorial away from being able to recreate the voice of whichever well-known person they choose. In addition, several new no-code or low-code options are on the market via a free and mobile-friendly app. Their availability has popularized spoofs of public figures uttering uncharacteristic and quirky lines across the internet.

Political parody is mostly harmless, but more pernicious cases are made possible with voice cloning. This is yet another tool malicious actors can use to peddle political disinformation or threaten security systems. Digital models of the unique vocal characteristics of an individual have previously been used for high security verification. Voice cloning technology complicates that system, making it more difficult for investigators – let alone casual media consumers – to distinguish between true and false audio recordings.
What is the difference between synthetic voice and voice cloning?

A synthetic voice is computer-generated speech created with artificial intelligence and deep learning. This is created using either text-to-speech (TTS) or speech-to-speech (STS) technology. Recent advancements in AI and deep learning have yielded synthetic voices that sound more human and less robotic to listeners. These improvements have led to the increased use of synthetic voices in advertising, entertainment, and education, as well as for chatbots and other virtual assistants.

Voice cloning is a sub-type of synthetic voice, wherein the user replicates a real person’s voice using AI. Voice cloning models are now able to mimic a specific person’s inflections and tone more accurately, where only a few years ago this was too costly. AI-enabled neural networks can now classify the main features of a voice, such as pronunciation, timbre, and rhythm, in great detail. These features constitute a “speaker profile” – the set of elements that distinguish one voice from another. Once equipped with a speaker profile, practitioners can train a model to recreate the target voice with TTS and STS. Because AI requires large amounts of existing audio recordings as training data to mimic a specific person’s inflections and tone with high accuracy, well-known people are naturally more at risk of being mimicked.

When paired with large-scale language models, such as GPT-3, it is relatively straightforward to automate the creation of online audio-based content at scale and speed, and empower malicious actors who previously lacked this technological capacity. For example, voice cloning could be used to make automated scam phone calls more sophisticated and widespread. More alarmingly, the automation of voice cloning makes the mass production of fake videos all the easier, and risks flooding the online information space with a slew of AI-generated video content. The sheer amount of fake content would make verification of every video more challenging, and the average person lacks the time to check. In such a scenario, fact-checkers could be overwhelmed by the sheer amount of artificial content, and users’ trust in what they view online would be eroded.
What are the risks of voice cloning technology becoming more accessible?

- Voice cloning can be used to create false news content or videos that appear to be legitimate, yet are designed to confuse and mislead viewers. When employed on a massive, automated scale, such videos can flood social media and reduce users’ abilities to distinguish between genuine and artificial content.
- Other public figures are also at risk, including celebrities. Voice clones could spread rumours about them, damage their reputations, or attribute to them false opinions, such as endorsements for certain political views or scams.
- Finally, as exhibited by the financial fraud committed against the earlier mentioned Japanese company, voice clones can facilitate cybercrime.

How can we respond to the misuse of voice cloning?

- AI companies should make an effort to increase personal protections in voice clone technology, including requiring a person’s consent by obtaining rights and permissions to use a speaker profile before a user can train their model on it.
- Cyber security firms should invest in the development of technology that recognizes AI-generated audio content, in addition to video, text, and image recognition technology.
- Social media companies need to adopt industry-wide standards for metadata tagging, as promoted by the Content Authenticity Initiative, so that their content moderator teams can more easily identify and flag AI-generated content.
- Government programmes geared towards protecting the public against phone scammers should adapt to this new threat, by educating the public about the increasing prevalence of voice cloning technology and what citizens should do if they receive an unusual request from a familiar-sounding voice.
Date: September 2023

This Rapid Response Brief was written by Waafa Heikal, Social Media Analyst at DRI’s MENA Regional Project with contributions by Duncan Allen, Digital Democracy Research Associate. The Brief is part of Democracy Reporting International’s Disinfo Radar project, funded by the German Federal Foreign Office. Its contents do not necessarily represent the position of the German Federal Foreign Office.

About Democracy Reporting International

DRI is an independent organisation dedicated to promoting democracy worldwide. We believe that people are active participants in public life, not subjects of their governments. Our work centres on analysis, reporting and capacity-building. For this, we are guided by the democratic and human rights obligations enshrined in international law. Headquartered in Berlin, DRI has offices in Lebanon, Libya, Myanmar, Pakistan, Sri Lanka, Tunisia, and Ukraine.